Igor Kortchemski
Probality theory,
Autumn 2023
ETH

# Chapter 7: convergence in distribution

Outline:



## 1) Definition and first properties

Motivation We have seen several notions of convergence of r.v $X_n \to X$ : a.s, in $\mathbb{P}$, $L^p$. In these cases, the quantity "$X_n(\omega) - X(\omega)$" was involved. Here we define a notion of convergence for the laws of r.v. For example, if $X_n$ and $X$ are $\mathbb{Z}$-valued, it is natural to say that their laws are "close" if $\forall k \in \mathbb{Z}$ $\mathbb{P}(X_n = k)$ is "close" to $\mathbb{P}(X=k)$.

However it is delicate to extend this to a general setting.

We will work in $\mathbb{R}^d$ ($d \geq 1$) but most of what follows can be extended to general metric spaces.

Notation $C_b(\mathbb{R}^d) = \{ f : \mathbb{R}^d \to \mathbb{R}$ continuous, bounded$\}$ and write $\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)|$ for $f \in C_b(\mathbb{R}^d)$. Here $|\cdot|$ denotes any norm on $\mathbb{R}^d$

Definition

- A sequence $(\mu_n)$ of probability measures is said to converge weakly to a probability measure $\mu$ on $\mathbb{R}^d$ if $\forall f \in C_b(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} f(x) \mu_n(dx) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}^d} f(x) \mu(dx)$

- A sequence $(X_n)$ of $\mathbb{R}^d$-valued r.v. is said to converge in distribution or to converge in law

to a $\mathbb{R}^d$-valued r.v. $X$ if $\mathbb{P}_{X_n} \to \mathbb{P}_X$ weakly, that is $\forall f \in C_b(\mathbb{R}^d)$, $\mathbb{E}[f(X_n)] \underset{n\to\infty}{\to} \mathbb{E}[f(X)]$
We write $X_n \xrightarrow{(d)} X$

**Remarks** • There is an abuse of language when we say "$X_n$ converges in distribution to $X$": for example, if $\mathbb{P}(X=1)=\mathbb{P}(X=-1)=\frac{1}{2}$, then $X \overset{law}{=} -X$, so setting $X_n = X$ we have $X_n \xrightarrow{(d)} X$ and $X_n \xrightarrow{(d)} -X$ ! The limiting r.v. is not uniquely defined: only its law is. For this reason, we sometimes say that "$X_n$ converges in distribution to $\mu$" (with $\mu$ probability measure)

• The r.v $(X_n)$, $X$ are not necessarily defined on the same probability space: convergence in distribution is VERY different from the a.s, $\mathbb{P}$, $L^p$ convergences.

# END OF LECTURE 22

**Examples**

• If $X_n$ is uniform on $\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$, then $X_n \xrightarrow{(d)}$ Uniform law on $[0,1]$. Indeed, if $f \in C_b(\mathbb{R})$,
$$\mathbb{E}[f(X_n)] = \sum_{k\geq 1}^{n} \frac{1}{n} f\left(\frac{k}{n}\right) \underset{n\to\infty}{\to} \int_0^1 f(x)\,dx \quad \text{(Rieman sum)}$$

• If $X_n$ is a $\mathcal{N}(0,\sigma_n^2)$ r.v with $\sigma_n \to 0$, then $X_n \xrightarrow{(d)} 0$ (constant r.v $=0$).
Indeed, $$\mathbb{E}[f(X_n)] = \int_{\mathbb{R}} f(x) \frac{e^{-\frac{x^2}{2\sigma_n^2}}}{\sqrt{2\pi\sigma_n^2}}\,dx = \int_{\mathbb{R}} f(x\sigma_n) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx \underset{n\to\infty}{\to} \int_{\mathbb{R}} f(0) \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx = f(0)$$ by dominated convergence

• If $\mu_n = \delta_{1/n}$, then $\mu_n \xrightarrow{weakly} \mu = \delta_0$ because $\forall f \in C_b(\mathbb{R})$, $\int_{\mathbb{R}} f(x)\,\mu_n(dx) = f(1/n) \underset{n\to\infty}{\to} f(0) = \int_{\mathbb{R}} f(x)\,\delta_0(dx)$
Observe that $0 = \mu_n(\{0\}) \underset{n\to\infty}{\not\to} \mu(\{0\}) = 1$
In particular $X_n \xrightarrow{(d)} X$ is NOT equivalent to $\forall B \in \mathcal{B}(\mathbb{R}^d)$, $\mu_n(B) \underset{n\to\infty}{\to} \mu(B)$

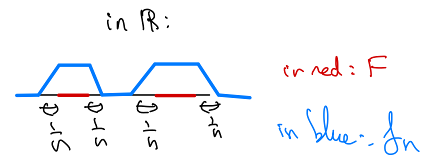
• If $X_n, X \in \mathbb{Z}$ we will later see that $X_n \xrightarrow{(d)} X$ iff $\forall k \in \mathbb{Z}$ $\mathbb{P}(X_n = k) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(X=k)$

**Lemma** If $X_n \xrightarrow{(d)} X$ and $X_n \xrightarrow{(d)} Y$ then $X$ and $Y$ have the same law (we write $X \overset{(d)}{=} Y$)

**Proof** By assumption $\forall f \in C_b(\mathbb{R}^d)$, $\mathbb{E}[f(X_n)] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[f(X)]$, so $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$
$\underset{n\to\infty}{\longrightarrow} \mathbb{E}[f(Y)]$

We show that $\forall A \in \mathcal{B}(\mathbb{R}^d)$, $\mathbb{P}_X(A) = \mathbb{P}_Y(A)$

**Step 1** We show that $\forall F \subset \mathbb{R}^d$ closed $\mathbb{P}_X(F) = \mathbb{P}_Y(F)$ by an approximation argument




Set $f_n(x) = \max(1 - n\, d(x,F), 0)$

Then $f_n \xrightarrow[n\to\infty]{} \mathbb{1}_F$ pointwise because $F$ is closed

In addition $f_n$ is continuous (exercise: check that $x \mapsto d(x,F)$ and $f_n$ are lipschitz), bounded by 1

So by dominated convergence
$$\begin{array}{ccc} \mathbb{E}[f_n(X)] & = & \mathbb{E}[f_n(Y)] \\ \downarrow{\scriptstyle n\to\infty} & & \downarrow{\scriptstyle n\to\infty} \\ \mathbb{P}_X(F) = \mathbb{P}(X\in F) = \mathbb{E}[\mathbb{1}_F(X)] & & \mathbb{E}[\mathbb{1}_F(Y)] = \mathbb{P}(Y\in F) = \mathbb{P}_Y(F) \end{array}$$

**Step 2** We know that $\mathcal{C} = \{A \in \mathcal{B}(\mathbb{R}^d) : \mathbb{P}_X(A) = \mathbb{P}_Y(A)\}$ is a Dynkin system, containing a generating $\pi$-system made of closed sets of $\mathbb{R}^d$. By the Dynkin lemma we conclude that $\mathcal{C} = \mathcal{B}(\mathbb{R}^d)$

$\sim$

**Proposition** If $X_n, X \in \mathbb{R}^d$ and $X_n \xrightarrow{(d)} X$, if $F : \mathbb{R}^d \to \mathbb{R}^n$ is continuous, then $F(X_n) \xrightarrow{(d)} F(X)$ (r.v in $\mathbb{R}^n$)

**Proof** Take $f : \mathbb{R}^n \to \mathbb{R}$ continuous bounded. Then $f\circ F : \mathbb{R}^d \to \mathbb{R}$ is continuous bounded, and $\mathbb{E}[f\circ F(X_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[f\circ F(X)]$, so $\mathbb{E}[f(F(X_n))] \xrightarrow[n\to\infty]{} \mathbb{E}[f(F(X))]$.

$\sim$

**Proposition** If $X_n, X \in \mathbb{R}^d$ and $X_n \xrightarrow{a.s} X$ or $X_n \xrightarrow{\mathbb{P}} X$ or $X_n \xrightarrow{L^p} X$, then $X_n \xrightarrow{(d)} X$.

**Proof** Since a.s. convergence and $L^p$ convergence implies convergence in probability, it suffices to show that $X_n \xrightarrow{\mathbb{P}} X$ implies $X_n \xrightarrow{(d)} X$. Argue by contradiction: let $f \in C_b(\mathbb{R}^d)$ be such that $\mathbb{E}[f(X_n)] \not\to \mathbb{E}[f(X)]$. Then there exist $\varepsilon > 0$ and a subsequence $\varphi$ s.t $\forall n \geq 1$, $|\mathbb{E}[f(X_{\varphi(n)})] - \mathbb{E}[f(X)]| \geq \varepsilon$ $(*)$. Since $X_{\varphi(n)} \xrightarrow{\mathbb{P}} X$, by the subsequence lemma, there exists a subsequence $\psi$ s.t $X_{\varphi\circ\psi(n)} \xrightarrow{a.s} X$. Then $f(X_{\varphi\circ\psi(n)}) \xrightarrow{a.s} f(X)$ by continuity, and since $f$ is bounded we get $\mathbb{E}[f(X_{\varphi\circ\psi(n)})] \to \mathbb{E}[f(X)]$ by dominated convergence, contradicting $(*)$.

$\sim$

Observe that we can have $X_n \xrightarrow{(d)} X$ and $\mathbb{E}[X_n] \underset{n\to\infty}{\not\to} \mathbb{E}[X]$ (take $X_n \geq 0$ with $X_n \xrightarrow{a.s} 0$ and $X_n \overset{L^1}{\not\to} 0$)
(consistent with the fact that $f(x) = x$ is not bounded)

## 2) The portemanteau / Alexandrov theorem

**Theorem (portemanteau / Alexandrov)** Let $(\mu_n), \mu$ be probability measures on $\mathbb{R}^d$. The following are $\Leftrightarrow$:

(1) $\mu_n \to \mu$ weakly

(2) $\forall f : \mathbb{R}^d \to \mathbb{R}$ lipschitz bounded, $\int_{\mathbb{R}^d} f(x)\, \mu_n(dx) \xrightarrow[n\to\infty]{} \int f(x)\, \mu(dx)$

(3) $\forall F \subset \mathbb{R}^d$ closed, $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$

(4) $\forall O \subset \mathbb{R}^d$ open, $\liminf_{n\to\infty} \mu_n(O) \geq \mu(O)$

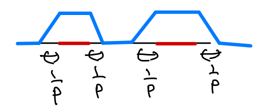
(5) $\forall B \in \mathcal{B}(\mathbb{R}^d)$ such that $\mu(\overline{B} \setminus \mathring{B}) = 0$, $\mu_n(B) \xrightarrow[n\to\infty]{} \mu(B)$

(6) $\forall f : \mathbb{R}^d \to \mathbb{R}$ measurable bounded, continuous at $\mu$-almost every point (i.e $\mu(\{x \in \mathbb{R}^d : f \text{ continuous at } x\}) = 1$)

we have $\int_{\mathbb{R}^n} f(x)\, \mu_n(dx) \xrightarrow[n\to\infty]{} \int f(x) \mu(dx)$

Proof $\boxed{(1) \Rightarrow (2)}$ ok because a Lipschitz function is continuous

$\boxed{(2) \Rightarrow (3)}$ Take $F$ closed and define as before $f_p(x) = \max(1 - p\,d(x,F), 0)$




in red: $F$

in blue: $f_p$

Then $0 \leq \mathbb{1}_F \leq f_p \leq 1$, $f_p \xrightarrow[p\to\infty]{} \mathbb{1}_F$ pointwise and $f_p$ is Lipschitz.

Thus $\mu_n(F) \leq \mu_n(f_p)$

so $\limsup_{n\to\infty} \mu_n(F) \leq \mu(f_p)$ and $\mu(f_p) \xrightarrow[p\to\infty]{} \mu(\mathbb{1}_F) = \mu(F)$ by dominated convergence

$\boxed{(3) \Rightarrow (4)}$ Apply (3) with $O^c$

$\boxed{(3)+(4) \Rightarrow (5)}$ Take $B \in \mathbb{R}^d$. Then

$$\mu(\mathring{B}) \leq \liminf_{n\to\infty} \mu_n(\mathring{B}) \leq \liminf_{n\to\infty} \mu_n(B) \leq \limsup_{n\to\infty} \mu_n(B) \leq \limsup_{n\to\infty} \mu_n(\overline{B}) \leq \mu(\overline{B})$$

If $\mu(\overline{B} \setminus \mathring{B}) = 0$ then $\mu(\mathring{B}) = \mu(B) = \mu(\overline{B})$ and all the previous $\leq$ are $=$

$\boxed{(5) \Rightarrow (6)}$ This is the delicate part. Let $f : \mathbb{R}^d \to \mathbb{R}$ be measurable bounded. Set $D = \{x \in \mathbb{R}^d : f \text{ not continuous at } x\}$ and assume $\mu(D) = 0$. By writing $f = f^+ - f^-$ we may assume $f \geq 0$. Let $K > 0$ be such that $0 \leq f \leq K$. By Fubini–Tonelli theorem, write

$$\int_{\mathbb{R}^d} f(x)\, \mu(dx) = \int_{\mathbb{R}^d} \left(\int_0^\infty \mathbb{1}_{t \leq f(x)}\, dt\right) \mu(dx) = \int_{\mathbb{R}^d} \left(\int_0^K \mathbb{1}_{t \leq f(x)}\, dt\right) \mu(dx) = \int_0^K \left(\int_{\mathbb{R}^d} \mathbb{1}_{t \leq f(x)}\, \mu(dx)\right) dt = \int_0^K \mu(A_t)\, dt$$

With $A_t = \{x \in \mathbb{R}^d : f(x) \geq t\}$

But $\overline{A_t} \setminus \mathring{A_t} \subset \{x \in \mathbb{R}^d : f(x) = t\} \cup D$. Indeed, if $x \in \overline{A_t}$ and $x \notin D$ then by continuity $f(x) \geq t$ and if $x \notin \mathring{A_t}$ then necessarily $f(x) = t$: indeed, otherwise $f(x) > t$ and $f(x') > t$ for $x'$ in a neighborhood of of $x$, contradicting $x \notin \mathring{A_t}$.

Also, $\{t \geq 0 : \mu(\{x \in \mathbb{R}^d : f(x) = t\}) > 0\}$ is at most countable (indeed, there are at most $k$ values of $t$ such that $\mu(\{x \in \mathbb{R} : f(x) = t\}) \geq \frac{1}{k}$.

As a consequence $\mu(\overline{A_t} \setminus \mathring{A_t}) = 0$ for almost all $t$ for the Lebesgue measure. By (4),

$\mu_n(A_t) \xrightarrow[n\to\infty]{} \mu(A_t)$ for almost all $t$ for the Lebesgue measure

Since $\mu_n(A_t) \leq 1$, by dominated convergence we get $\int_0^K \mu_n(A_t)\,dt \xrightarrow[n\to\infty]{} \int_0^K \mu(A_t)\,dt = \int_{\mathbb{R}^d} f(x)\,\mu(dx)$.

(6) ⇒ (1) ok because $f$ continuous implies $f$ continuous at $\mu$-almost every point

**END OF LECTURE 23 ~ ((5) ⇒ (6) not done in class)**

<u>Remark</u> It is possible to check that $\{x \in \mathbb{R}^d : f \text{ is continuous at } x\} \in \mathcal{B}(\mathbb{R}^d)$ for $f$ measurable.

<u>Probabilistic reformulation</u> Let $X_n, X$ be r.v in $\mathbb{R}^d$. Then the following are ⇔

① $X_n \xrightarrow{(d)} X$

② $\forall f : \mathbb{R}^d \to \mathbb{R}$ lipschitz bounded $\mathbb{E}[f(X_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[f(X)]$

③ $\forall F \subset \mathbb{R}^d$ closed, $\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$

④ $\forall O \subset \mathbb{R}^d$ open, $\liminf_{n\to\infty} \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)$

⑤ $\forall B \in \mathbb{R}^d$, $\mathbb{P}(X \in \bar{B} \setminus \mathring{B}) = 0 \Rightarrow \mathbb{P}(X_n \in B) \xrightarrow[n\to\infty]{} \mathbb{P}(X \in B)$

⑥ $\forall f : \mathbb{R}^d \to \mathbb{R}$ measurable bounded, a.s. continuous at $X$, then $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$

We now give some applications

<u>Corollary</u> (Extended continuous mapping) If $X_n \xrightarrow{(d)} X$ and $F : \mathbb{R}^d \to \mathbb{R}^n$ is measurable, a.s. continuous at $X$, then $F(X_n) \xrightarrow{(d)} F(X)$

This simply comes from ⑥: if $f : \mathbb{R}^n \to \mathbb{R}$ is continuous bounded, then $f \circ F$ is measurable, bounded, a.s. continuous at $X$

<u>Example</u> In $\mathbb{R}$, if $X_n \xrightarrow{(d)} X$ and $X \neq 0$ a.s, then $\frac{1}{X_n} \xrightarrow[n\to\infty]{(d)} \frac{1}{X}$, with the convention $\frac{1}{x} = 0$ when $x = 0$.

If $X$ is $\mathbb{R}$ valued, recall the notation $F_X(t) = \mathbb{P}(X \leq t)$, $t \in \mathbb{R}$, for its cdf. We have seen that:

- $F_X$ is continuous at $t$ iff $\mathbb{P}(X = t) = 0$
- The number of discontinuity points of $F_X$ is at most countable; in particular its points of continuity are dense in $\mathbb{R}$.

<u>Theorem</u> Let $(X_n), X$ be $\mathbb{R}$ valued r.v

$X_n \xrightarrow{(d)} X$ iff $F_{X_n}(t) \xrightarrow[n\to\infty]{} F_X(t)$ for every $t$ continuity point of $F_X$

**Proof** [⇒] Let $t \in \mathbb{R}$ be a continuity point of $F_X$. We apply portmanteau with $B = (-\infty, t]$: $\bar{B} \setminus \mathring{B} = \{t\}$ and $\mathbb{P}(X = t) = 0$ because $F_X$ is continuous at $t$. So $F_{X_n}(t) = \mathbb{P}(X_n \leq t) \xrightarrow[n\to\infty]{} \mathbb{P}(X \leq t) = F_X(t)$.

[⇐] We show that $\forall\, O \subset \mathbb{R}$ open, $\liminf_{n\to\infty} \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)$ $(*)$. This implies $X_n \xrightarrow{(d)} X$ by Portmanteau.

**Step 1**: $O$ is an open interval. We show that

$\limsup_{n\to\infty} \mathbb{P}(X_n \leq a) \overset{(*)}{\leq} \mathbb{P}(X \leq a)$ and $\liminf_{n\to\infty} \mathbb{P}(X_n < b) \overset{(**)}{\geq} \mathbb{P}(X < b)$ for $a, b \in \mathbb{R}$, which implies $(*)$ with $O = (a, b)$, $a < b$, $a, b \in \mathbb{R} \cup \{\pm\infty\}$. Indeed, writing $\mathbb{P}(a < X_n < b) = \mathbb{P}(X_n > a) - \mathbb{P}(X_n \geq b)$, we get

$$\liminf_{n\to\infty} \mathbb{P}(a < X_n < b) = \liminf_{n\to\infty} \mathbb{P}(X_n < b) - \limsup_{n\to\infty} \mathbb{P}(X_n \leq a) \geq \mathbb{P}(X < b) - \mathbb{P}(X \leq a) = \mathbb{P}(a < X < b)$$

- For $(*)$, take $x > a$ with $F_X$ continuous at $x$. Then $\limsup_{n\to\infty} \mathbb{P}(X_n \leq a) \leq \limsup_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$. Now take $x \downarrow a$ (possible because $F_X$ is continuous on a dense set). By right continuity $\mathbb{P}(X \leq x) \xrightarrow[x\downarrow a]{} \mathbb{P}(X \leq a)$ and we get $(*)$.
- Similarly, for $(**)$, take $x < b$ with $F_X$ continuous at $x$. We have $\liminf_{n\to\infty} \mathbb{P}(X_n < b) \geq \liminf_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$. Now take $x \uparrow b$. We have $\mathbb{P}(X \leq x) \nearrow \mathbb{P}(X < b)$ and $(**)$ follows.

**Step 2** $O$ open. We know that we can write $O = \bigcup_{i \in I} (a_i, b_i)$ as a countable disjoint union of open intervals. Then

$$\mathbb{P}(X \in O) = \sum_{i \in I} \mathbb{P}(X \in (a_i, b_i)) \leq \sum_{i \in I} \liminf_{n\to\infty} \mathbb{P}(X_n \in (a_i, b_i)) \leq \liminf_{n\to\infty} \sum_{i \in I} \mathbb{P}(X_n \in (a_i, b_i)) \quad \text{(Fatou)}$$
$$= \liminf_{n\to\infty} \mathbb{P}(X_n \in O)$$

**Corollary** If $F_X$ is continuous (i.e. $\forall t \in \mathbb{R}$ $\mathbb{P}(X = t) = 0$: $X$ has no atoms), then $X_n \xrightarrow{(d)} X$ iff $F_{X_n} \to F_X$ pointwise iff $\mathbb{P}(X_n < t) \xrightarrow[n\to\infty]{} \mathbb{P}(X < t)$ (see exercise sheet).

In particular $X_n \xrightarrow{(d)}$ r.v with density $p$ $\Leftrightarrow$ $\forall a, b \in \mathbb{R}$ $\mathbb{P}(a \leq X_n \leq b) \xrightarrow[n\to\infty]{} \int_a^b p(s)\, ds$.

here $\leq$ or $<$    here $\leq$ or $<$

**Application** Fix $\lambda > 0$ and $(X_n)_{n \geq 1}$ a sequence of r.v with $X_n \sim \text{Geo}(\frac{\lambda}{n})$. Then $\frac{X_n}{n} \xrightarrow{(d)} \text{Exp}(\lambda)$.

**Proof**: Let $X$ be a $\text{Exp}(\lambda)$ r.v.. Since $X$ has no atoms, it is enough to show that $\forall t \geq 0$, $\mathbb{P}(\frac{X_n}{n} \geq t) \xrightarrow[n\to\infty]{} \mathbb{P}(X \geq t)$ (for $t < 0$, they are 0).

Write $\mathbb{P}(X_n \geq nt) = \mathbb{P}(X_n \geq \lceil nt \rceil) = \left(1 - \frac{\lambda}{n}\right)^{\lceil nt \rceil}$

$$= \exp\left(\lceil nt \rceil \ln\left(1 - \frac{\lambda}{n}\right)\right) = \exp\left((nt + O(1))\left(-\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)\right) = \exp(-\lambda t + o(1)) \xrightarrow[n\to\infty]{} \exp(-\lambda t).$$
$$= \mathbb{P}(X \geq t)$$

**Proposition** Let $(X_n)_{n\geq 1}$ be $\mathbb{R}^d$-valued r.v and $a \in \mathbb{R}^d$ a constant. Then $X_n \xrightarrow{(d)} a$ iff $X_n \xrightarrow{\mathbb{P}} a$

Proof: $\boxed{\Leftarrow}$ We know that convergence in probability implies convergence in distribution

$\boxed{\Rightarrow}$ Set $\varepsilon > 0$ and let $B(a,\varepsilon) = \{x \in \mathbb{R}^d : |x-a| < \varepsilon\}$ be an open ball centered at $a$

Then by Portemanteau $\limsup_{n\to\infty} \mathbb{P}(|X_n - a| \geq \varepsilon) = \limsup_{n\to\infty} \mathbb{P}(X_n \in \underbrace{B(a,\varepsilon)^c}_{\text{closed}}) \leq \mathbb{P}(a \in B(a,\varepsilon)^c) = 0$

**Proposition** Let $(X_n), (Y_n), X$ be $\mathbb{R}^d$-valued r.v. Assume that $X_n \xrightarrow{(d)} X$, $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0$. Then $Y_n \xrightarrow{(d)} X$

Proof: We show that $\forall F \subset \mathbb{R}^d$ closed, $\limsup_{n\to\infty} \mathbb{P}(Y_n \in F) \leq \mathbb{P}(X \in F)$, and the result will follow by Porte-manteau

To do this, fix $p \geq 1$ and set $F^{1/p} = \{x \in \mathbb{R}^d : d(x,F) \leq \frac{1}{p}\}$ the $\frac{1}{p}$-closed neighborhood of $F$ (exercise: check that it is closed)

Then $\mathbb{P}(Y_n \in F) = \mathbb{P}(Y_n \in F, |X_n - Y_n| \leq \frac{1}{p}) + \mathbb{P}(Y_n \in F, |X_n - Y_n| > \frac{1}{p})$

$\leq \mathbb{P}(X_n \in F^{1/p}) + \mathbb{P}(|X_n - Y_n| > \frac{1}{p})$

So $\limsup_{n\to\infty} \mathbb{P}(Y_n \in F) \leq \mathbb{P}(X \in F^{1/p}) + 0$ by assumption

But $\mathbb{P}(X \in F^{1/p}) \xrightarrow[p\to\infty]{} \mathbb{P}(X \in F)$ because $F^{1/p}$ is decreasing in $p$ and $\bigcap_{p\geq 1} F^{1/p} = F$ because $F$ is closed.

We conclude that $\limsup_{n\to\infty} \mathbb{P}(Y_n \in F) \leq \mathbb{P}(X \in F)$.

Important application:

**Theorem (Slutsky)** Let $(X_n), (Y_n), X$ be $\mathbb{R}^d$-r.v. Let $a \in \mathbb{R}^d$ be a constant. Assume that $X_n \xrightarrow{(d)} X$ and $Y_n \xrightarrow{\mathbb{P}} a$

Then $(X_n, Y_n) \xrightarrow{(d)} (X, c)$

Proof: • $(X_n, a) \xrightarrow{(d)} (X, a)$ by continuity of $f(x) = (x, c)$.

• $|(X_n, a) - (X_n, Y_n)| = |a - Y_n| \xrightarrow{\mathbb{P}} 0$ (here $|(x,y)| = |x| + |y|$)

Hence the result by the previous proposition

**Example** in $\mathbb{R}$, if $X_n \xrightarrow{(d)} X$ and $Y_n \xrightarrow{\mathbb{P}} a \neq 0$, then $\frac{X_n}{Y_n} \xrightarrow{(d)} \frac{X}{a}$.

(Indeed, $(X_n, Y_n) \xrightarrow{(d)} (X, a)$ and $f(x,y) = \begin{cases} \frac{x}{y} & \text{if } y \neq 0 \\ 0 & \text{if } y = 0 \end{cases}$ is a.s continuous at $(X, a)$, so $f(X_n, Y_n) \xrightarrow{(d)} f(X, a)$

**Example** in $\mathbb{R}^d$, if $X_n \xrightarrow{(d)} X$ and $Y_n \xrightarrow{\mathbb{P}} 0$, then $X_n + Y_n \xrightarrow{(d)} X$

(indeed, $(X_n, Y_n) \xrightarrow{(d)} (X, 0)$ and $f(x,y) = x + y$ is continuous, so $f(X_n, Y_n) \xrightarrow{(d)} f(X, 0)$)

⚠ ⚠ In general, $X_n \xrightarrow{(d)} X$, $Y_n \xrightarrow{(d)} Y$ does NOT imply $(X_n, Y_n) \xrightarrow{(d)} (X,Y)$.

Indeed, take $X$ with $\mathbb{P}(X=1) = \mathbb{P}(X=-1) = \frac{1}{2}$, $X_n = X$, $Y_n = -X$. Then $X_n \xrightarrow{(d)} X$, $Y_n \xrightarrow{(d)} X$ but $\underset{(X,-X)}{\underset{\parallel}{(X_n, Y_n)}} \overset{(d)}{\not\longrightarrow} (X,X)$

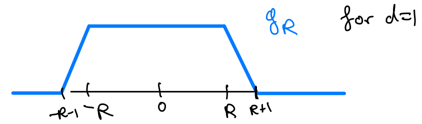
But we will later see that the result is true if $X_n \perp\!\!\!\perp Y_n$

## 3) Restricting the class of test functions

Set $\mathcal{C}_c(\mathbb{R}^d) = \{ f : \mathbb{R}^d \to \mathbb{R}$ continuous with compact support $\}$

**Proposition** Let $\mu_n, \mu$ be probability measures on $\mathbb{R}^d$. Then $\mu_n \to \mu$ weakly iff $\forall f \in \mathcal{C}_c(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} f(x)\mu_n(dx) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}^d} f(x)\mu(dx)$

**Proof** $\boxed{\Rightarrow}$ Clear because $\mathcal{C}_c(\mathbb{R}^d) \subset \mathcal{C}_b(\mathbb{R}^d)$




$\boxed{\Leftarrow}$ Take $f \in \mathcal{C}_b(\mathbb{R}^d)$. We show $\int_{\mathbb{R}^d} f(x)\mu_n(dx) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}^d} f(x)\mu(dx)$

We use a truncation argument: for $R>0$, set $g_R(x) = \begin{cases} 1 & \text{if } |x| \le R \\ \max(R+1-|x|, 0) & \text{if } |x| > R \end{cases}$

For $R>0$ fixed, $\left| \int f(x)\mu_n(dx) - \int f(x)\mu(dx) \right| \le \int |f(x) - f(x)g_R(x)| \mu_n(dx) \le \|f\|_\infty \left(1 - \int g_R(x)\mu_n(dx)\right)$ $\xrightarrow[n\to\infty]{}$ $\int g_R(x)\mu(dx)$ because $g_R \in \mathcal{C}_c(\mathbb{R}^d)$

$+ \left| \int f(x) g_R(x) \mu_n(dx) - \int f(x) g_R(x) \mu(dx) \right| \xrightarrow[n\to\infty]{} 0$ because $f g_R \in \mathcal{C}_c(\mathbb{R}^d)$

$+ \int |f(x) - f(x) g_R(x)| \mu(dx) \le \|f\|_\infty \left(1 - \int g_R(x)\mu(dx)\right)$

So $\limsup_{n\to\infty} \left| \int f(x)\mu_n(dx) - f(x)\mu(dx) \right| \le 2\|f\|_\infty \left(1 - \int g_R(x)\mu(dx)\right) \xrightarrow[R\to\infty]{} 0$ by monotone convergence, because $g_R \uparrow 1$.

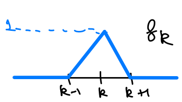
$\sim$

⚠ Contrary to previous statements, this result does not extend to general metric spaces

# END OF LECTURE 24

**Corollary** Let $X_n, X$ be $\mathbb{Z}$-valued r.v. Then $X_n \xrightarrow{(d)} X$ iff $\forall k \in \mathbb{Z}$, $\mathbb{P}(X_n = k) \xrightarrow[n\to\infty]{} \mathbb{P}(X=k)$

**Proof**: $\boxed{\Rightarrow}$ Fix $k \in \mathbb{Z}$




$g_k \in \mathcal{C}_c(\mathbb{R})$, and $\mathbb{P}(X_n = k) = \mathbb{E}[g_k(X_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[g_k(X)] = \mathbb{P}(X=k)$

$\boxed{\Leftarrow}$ Take $f \in \mathcal{C}_c(\mathbb{R})$. We show $\mathbb{E}[f(X_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[f(X)]$. Assume that $[-K,K] \subset \text{support} f$, so that $\text{Card}([-K,K] \cap \mathbb{Z}) < \infty$

Then $\mathbb{E}[f(X_n)] = \sum_{k \in [-K,K] \cap \mathbb{Z}} \mathbb{P}(X_n = k) f(k) \xrightarrow[n\to\infty]{} \sum_{k \in [-K,K] \cap \mathbb{Z}} \mathbb{P}(X=k) f(k) = \mathbb{E}[f(X)]$

$\sim$

**Application** Fix $\lambda>0$. Let $(X_n)$ be r.v with $X_n \sim Bin(n, \frac{\lambda}{n})$. Then $X_n \xrightarrow{(d)} Poi(\lambda)$.

**Proof**: We show that $\forall k \geq 0$, $\mathbb{P}(X_n = k) \xrightarrow[n\to\infty]{} e^{-\lambda} \cdot \frac{\lambda^k}{k!}$

We have
$$\mathbb{P}(X_n=k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$
$$= \underbrace{\frac{n(n-1)\cdots(n-k+1)}{n^k}}_{\xrightarrow[n\to\infty]{} 1} \times \frac{\lambda^k}{k!} \exp\Big(\underbrace{(n-k)\ln\left(1-\frac{\lambda}{n}\right)}_{\xrightarrow[n\to\infty]{} -\lambda}\Big) \xrightarrow[n\to\infty]{} e^{-\lambda}\frac{\lambda^k}{k!}$$

# 4) Characteristic functions and Lévy's theorem

Characteristic functions are very useful to study laws and convergence in distribution in $\mathbb{R}^d$. They are expectations of complex-valued random variables. For the moment we have only considered integrals of $\mathbb{R}$-valued r.v. If $Z$ is a $\mathbb{C}$-valued r.v, when $\mathbb{E}[|Z|]<\infty$ (with $|x+iy| = \sqrt{x^2+y^2}$), we define $\mathbb{E}[Z] = \mathbb{E}[Re(Z)] + i\,\mathbb{E}[Im(Z)]$, which makes sense since $|Re(z)| \leq |z|$ and $|Im(Z)| \leq |Z|$, so $Re\, Z$ and $Im\, Z$ are integrable.

**Definition** The characteristic function of a $\mathbb{R}^d$-valued r.v $X$ is the function $\varphi_X : \mathbb{R}^d \to \mathbb{C}$ defined by
$$\varphi_X(u) = \mathbb{E}[e^{i\langle u, X\rangle}] = \mathbb{E}[e^{i(u_1X_1 + \cdots + u_nX_n)}] \text{ for } u \in \mathbb{R}^d.$$

It is well defined since $e^{i\langle u, X\rangle}$ is bounded and thus in $L^1$!

**Remark** By the transfer theorem, $\varphi_X(u) = \int_{\mathbb{R}^d} e^{i\langle u,x\rangle}\, \mathbb{P}_X(dx)$ is the Fourier transform of the probability measure $\mathbb{P}_X$

**Example** Take $X \sim Poi(\lambda)$. Then $\varphi_X(u) = \mathbb{E}[e^{iuX}] = \sum_{k\geq 0}^{\infty} \mathbb{P}(X=k)e^{iuk} = \sum_{k\geq 0}^{\infty} e^{-\lambda} \frac{\overbrace{\lambda^k e^{iuk}}^{=(\lambda e^{iu})^k}}{k!} = e^{\lambda e^{iu} - \lambda}$.

**Proposition** $\varphi_X$ satisfies:

1. $\varphi_X(0) = 1$
2. $\varphi_X(-u) = \overline{\varphi_X(u)} \quad \forall u \in \mathbb{R}^d$
3. $|\varphi_X(u)| \leq 1 \quad \forall u \in \mathbb{R}^d$
4. $|\varphi_X(u+h) - \varphi_X(u)| \leq \mathbb{E}[|e^{i\langle h,X\rangle} - 1|] \quad \forall u,h \in \mathbb{R}^d$, so $\varphi_X$ is uniformly continuous by dominated convergence

These properties immediately follow from the definition.

**Example** Take $X \sim N(m,\sigma^2)$; recall that $X$ has density $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$. Then $\varphi_X(u) = e^{ium - \frac{\sigma^2 u^2}{2}}$ for $u \in \mathbb{R}$.

Indeed, since $X - m \overset{(d)}{=} \sigma X'$ with $X' \sim N(0,1)$, it is enough to consider the case $m=0$ and $\sigma = 1$.

Since $X \overset{(d)}{=} -X$, $\varphi_X(u) = \varphi_X(-u)$, so $\varphi_X(u) \in \mathbb{R}$. In particular for $u \in \mathbb{R}$

$$\varphi_X(u) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{iux} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \cos(ux) e^{-x^2/2} dx.$$

To compute this integral, observe that $\varphi_X'(u) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} -x \sin(ux) e^{-x^2/2} dx$. Indeed, $|x \sin(ux) e^{-x^2/2}| \leq |x| e^{-x^2/2}$ is integrable (we use the theorem of derivation of integral with parameters).

Thus by integration by parts $\varphi_X'(u) = -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} u \cos(ux) dx = -u \varphi_X(u)$.

Solving this differential equation with $\varphi_X(0)=1$ gives $\varphi_X(u) = e^{-u^2/2}$.

**Remark** We often use the fact that if $X \sim N(m,\sigma^2)$ then $a + bX \sim N(a+m, b^2\sigma^2)$ for $a, b \in \mathbb{R}$

An important feature is that characteristic functions characterize laws.

**Theorem** Let $X, Y$ be $\mathbb{R}^d$-valued random variables. They have same law iff $\varphi_X = \varphi_Y$ (that is $\varphi_X(u) = \varphi_Y(u)$ $\forall u \in \mathbb{R}^d$)

**Proof** We do the proof in the case $d=1$ (the general case is similar).

$\boxed{\Rightarrow}$ Fix $u \in \mathbb{R}$. Since $x \mapsto e^{itx}$ is measurable bounded ($|e^{itx}| \leq 1$), we have by the transfer theorem

$$\mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} \mathbb{P}_X(dx) = \int_{\mathbb{R}} e^{itx} \mathbb{P}_Y(dx) = \mathbb{E}[e^{itY}]$$

$\boxed{\Leftarrow}$ Assume that $\varphi_X = \varphi_Y$. The idea is to add a small Gaussian perturbation: let $Z_n$ be a $N(0, \frac{1}{n^2})$ r.v., $\perp\!\!\!\perp$ of $X, Y$ (we may assume that $X, Y, Z_n$ are defined on a same probability space, see the remark after the proof)

we show that $X + Z_n \overset{(d)}{=} Y + Z_n$. $(\star)$

Indeed, assuming $(\star)$, since $\mathbb{E}[Z_n^2] = \frac{1}{n^2} \xrightarrow[n\to\infty]{} 0$, $Z_n \xrightarrow{L^2} 0$, so $Z_n \xrightarrow{\mathbb{P}} 0$, so $X + Z_n \xrightarrow{(d)} X$ and similarly $Y + Z_n \xrightarrow{(d)} Y$. By uniqueness in law of the limit, we get $X \overset{(d)}{=} Y$.

To show $(\star)$ we show that for $F \geq 0$ measurable, $\mathbb{E}[F(X+Z_n)]$ only depends on $\varphi_X$.

Set $g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$ and observe that by the above example $g_\sigma(z) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{iuz} g_{1/\sigma}(u) du$ for $z \in \mathbb{R}$

By the transfer theorem and Fubini-Tonelli

$$\mathbb{E}[F(X+Z_n)] = \mathbb{E}\Big[\int_{\mathbb{R}} g_{1/n}(z)\,dz\, F(X+z)\Big] = \mathbb{E}\Big[\int_{\mathbb{R}} g_{1/n}(z-X)\,F(z)\,dz\Big] = \int_{\mathbb{R}} dz\, F(z)\,\mathbb{E}[g_{1/n}(z-X)]$$

But

$$g_{1/n}(z-X) = \frac{n}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iu(z-X)} g_n(u)\,du,$$ so by Fubini-Lebesgue we get:

$$\mathbb{E}[g_{1/n}(z-X)] = \frac{n}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iuz} g_n(u)\,\mathbb{E}[e^{-iuX}]\,du.$$

We conclude that $$\mathbb{E}[F(X+Z_n)] = \int_{\mathbb{R}} dz\, F(z) \left(\frac{n}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iuz} g_n(u)\,\varphi_X(-u)\,du\right)$$

Thus $\varphi_X = \varphi_Y$ implies $\mathbb{E}[F(X+Z_n)] = \mathbb{E}[F(Y+Z_n)]$ and completes the proof.

Remark Let us justify the fact that if $X, Y$ are random variables (not necessarily defined on a same probability space) we can define $X', Y'$ on a same probability space with $X \overset{(d)}{=} X'$, $Y \overset{(d)}{=} Y'$ and $X' \perp\!\!\!\perp Y'$!

Assume that $X$ is defined on $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $Y$ on $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$.

Set $\Omega' = \Omega_1 \times \Omega_2$, $\mathcal{A}' = \mathcal{F}_1 \otimes \mathcal{F}_2$ (product $\sigma$-field), $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$ (product probability measure, and for $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$, $X'(\omega_1, \omega_2) = X(\omega_1)$ and $Y'(\omega_1, \omega_2) = Y(\omega_2)$.

# END OF LECTURE 25

Important corollary $\mathbb{R}$-valued r.v $X_1, \dots, X_k$ are $\perp\!\!\!\perp$ iff $\forall\, u_1, \dots, u_k \in \mathbb{R}$:

$$\mathbb{E}[e^{iu_1X_1 + \dots + iu_kX_k}] = \mathbb{E}[e^{iu_1X_1}] \cdots \mathbb{E}[e^{iu_kX_k}]$$

Proof $\boxed{\Rightarrow}$ When $X_1, \dots, X_k$ are $\perp\!\!\!\perp$, $\mathbb{E}[f_1(X_1) \cdots f_k(X_k)] = \mathbb{E}[f_1(X_1)] \cdots \mathbb{E}[f_k(X_k)]$ when $f_1(X_1), \dots, f_k(X_k)$ are integrable

$\boxed{\Leftarrow}$ Then $(X_1, \dots, X_k)$ has the same characteristic function as a r.v with law $\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_k}$: by the theorem, $\mathbb{P}_{(X_1, \dots, X_k)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_k}$, so $X_1, \dots, X_k$ are $\perp\!\!\!\perp$.

Important application Assume that $X \overset{(d)}{=} \mathcal{N}(m_1, \sigma_1^2)$ and $Y \overset{(d)}{=} \mathcal{N}(m_2, \sigma_2^2)$ and $X \perp\!\!\!\perp Y$. Then $X+Y \overset{(d)}{=} \mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$

Proof: We compute $\varphi_{X+Y}$: $\mathbb{E}[e^{iu(X+Y)}] = \mathbb{E}[e^{iuX}]\,\mathbb{E}[e^{iuY}]$ by $\perp\!\!\!\perp$

$$= e^{ium_1 - \frac{u^2\sigma_1^2}{2}} \cdot e^{ium_2 - \frac{u^2\sigma_2^2}{2}}$$

$$= e^{iu(m_1+m_2) - \frac{u^2}{2}(\sigma_1^2+\sigma_2^2)},$$

which is the characteristic function of a $N(m_1+m_2, \sigma_1^2+\sigma_2^2)$ r.v.

⚠ The result is false without the independence assumption: if $X \sim N(0,1)$, then $-X \sim N(0,1)$ but $0 = X - X \not\sim N(0,2)$.

Characteristic functions also allow to prove convergence in distribution

**Theorem (Lévy)** Let $X_n, X$ be $\mathbb{R}^d$-valued r.v. Then $X_n \xrightarrow{(d)} X$ iff $\forall u \in \mathbb{R}^d$, $\varphi_{X_n}(u) \underset{n\to\infty}{\longrightarrow} \varphi_X(u)$.

Proof To simplify notation we do the proof for $d=1$ (the general case is similar)

$\boxed{\Rightarrow}$ for $u \in \mathbb{R}$, $x \mapsto e^{ixu}$ is continuous bounded, so $\mathbb{E}[e^{iX_n u}] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[e^{iXu}]$ by definition of convergence in distribution.

$\boxed{\Leftarrow}$ We use the same idea of a small Gaussian perturbation. Let $Z_k \sim N(0, \frac{1}{k^2})$

Step 1: for every $k \geq 1$, $X_n + Z_k \xrightarrow{(d)} X + Z_k$.

Indeed, take $F: \mathbb{R} \to \mathbb{R}$ continuous with compact support, and recall that

$$\mathbb{E}[F(X_n+Z_k)] = \int_{\mathbb{R}} dz\, F(z) \left( \frac{k}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iuz} g_k(u)\, \varphi_{X_n}(-u)\, du \right)$$

Then

- $e^{iuz} g_k(u)\, \varphi_{X_n}(-u) \underset{n\to\infty}{\longrightarrow} e^{iuz} g_k(u)\, \varphi_X(-u)$ by assumption, and $|e^{iuz} g_k(u)\, \varphi_{X_n}(-u)| \leq g_k(u)$, integrable
- Thus $\frac{k}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iuz} g_k(u)\, \varphi_{X_n}(-u)\, du \underset{n\to\infty}{\longrightarrow} \frac{k}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iuz} g_k(u)\, \varphi_X(-u)\, du$ by dominated convergence

In addition $\left| \underbrace{\quad}_{\downarrow} \right| \leq \frac{k}{\sqrt{2\pi}} \int_{\mathbb{R}} g_k(u)\, du = 1$, and $\int_{\mathbb{R}} dz\, F(z)\, 1 < \infty$ because $F$ has compact support.

We conclude by dominated convergence again

$$\mathbb{E}[F(X_n + Z_k)] \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} dz\, F(z) \left( \frac{k}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iuz} g_k(u)\, \varphi_X(-u)\, du \right) = \mathbb{E}[F(X+Z_k)]$$

Step 2 We show $X_n \xrightarrow{(d)} X$. Take $F: \mathbb{R} \to \mathbb{R}$ $L$-lipschitz, bounded and write for $k \geq 1$: $|\mathbb{E}[F(X_n)] - \mathbb{E}[F(X)]|$

$$\leq \mathbb{E}[|F(X_n) - F(X_n + Z_k)|] + |\mathbb{E}[F(X_n + Z_k)] - \mathbb{E}[F(X + Z_k)]| + \mathbb{E}[|F(X+Z_k) - F(X)|]$$

$$\leq 2L \underbrace{\mathbb{E}[|Z_k|]}_{= \frac{1}{k}\mathbb{E}[|Z_1|]} + |\mathbb{E}[F(X_n + Z_k) - \mathbb{E}[F(X+Z_k)]]$$

because $Z_k \overset{(d)}{=} \frac{1}{k} Z_1$

Thus $\limsup_{n\to\infty} |\mathbb{E}[F(X_n)] - \mathbb{E}[F(X)]| \leq \frac{2L}{k} \mathbb{E}[|Z_1|]$ by Step 1.

By taking $k \to \infty$ we get the result.

∾

Application Assume that $X_n \xrightarrow{(d)} X$, $Y_n \xrightarrow{(d)} Y$ with $X_n \perp\!\!\!\perp Y_n$. Then $(X_n, Y_n) \xrightarrow{(d)} (X, Y)$ with $X, Y$ $\perp\!\!\!\perp$.

(equivalently, if $\mathbb{P}_{X_n} \to \mathbb{P}_X$ and $\mathbb{P}_{Y_n} \to \mathbb{P}_Y$, then $\mathbb{P}_{X_n} \otimes \mathbb{P}_{Y_n} \xrightarrow[n\to\infty]{} \mathbb{P}_X \otimes \mathbb{P}_Y$ where convergences are weak convergences of probability measures)

Proof: Write $\mathbb{E}[e^{iuX_n + ivY_n}] = \mathbb{E}[e^{iuX_n}]\,\mathbb{E}[e^{ivY_n}] \xrightarrow[n\to\infty]{} \mathbb{E}[e^{iuX}]\,\mathbb{E}[e^{ivX}] = \mathbb{E}[e^{iuX + ivY}]$

Then $\varphi_{(X_n, Y_n)} \xrightarrow{\text{pointwise}} \varphi_{(X,Y)}$ and the result follows.

∾

# 5) The Central limit theorem

Theorem Let $(X_n)_{n\geq 1}$ be an iid sequence of real-random variables with $\mathbb{E}[X_1^2] < \infty$. Set $\sigma^2 = \mathrm{Var}(X_1)$. Assume that $\sigma^2 > 0$. Then $\dfrac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} \xrightarrow{(d)} N(0,1)$

Remarks
- $\mathbb{E}[X_1^2] < \infty \Rightarrow X_1$ is integrable
- $\sigma^2 > 0$ rules out constant random variables
- equivalently, $\dfrac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sqrt{n}} \xrightarrow[n\to\infty]{(d)} N(0, \sigma^2)$ (because $\sigma \cdot N(0,1) \overset{(d)}{=} N(0, \sigma^2)$)
- equivalently, $\forall a < b$, $\mathbb{P}\left(a < \dfrac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} < b\right) \xrightarrow[n\to\infty]{} \displaystyle\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ (because the cdf of $N(0,1)$ is continuous)

- By the strong law of large numbers, $\frac{X_1+\dots+X_n}{n} - \mathbb{E}[X_1] \xrightarrow{a.s.} 0$. The Central Limit theorem says that the "speed of convergence" is of order $\frac{1}{\sqrt{n}}$.

The proof is based on the following estimate:

**Lemma** Assume that $X$ is $\mathbb{R}$-valued and $\mathbb{E}[X^2]<\infty$. Then $\Phi_X(t)=\mathbb{E}[e^{iXt}] = 1 + i\,\mathbb{E}[X]t - \frac{1}{2}\mathbb{E}[X^2]t^2 + o(t^2)$.

*Proof* We show that $\Phi_X$ is twice differentiable with $\Phi'_X(0)= i\,\mathbb{E}[X]$ and $\Phi''_X(0) = -\frac{1}{2}\mathbb{E}[X^2]$. This follows from a general theorem from measure theory (permutating differentiation and integration):

if $\begin{cases} \forall t\in\mathbb{R}, \quad F(t,X)\in L^1 \\ \text{a.s.} \quad t\mapsto F(t,X) \text{ is differentiable} \\ \exists Y\in L^1 \text{ st } \forall t\in\mathbb{R}, \ \left|\frac{\partial}{\partial t}F(t,X)\right|\leq Y \end{cases}$ then $t\mapsto \mathbb{E}[F(t,X)]$ is differentiable and $\frac{d}{dt}\mathbb{E}[F(t,X)] = \mathbb{E}[\frac{d}{dt}F(t,X)]$

Here $F(t,X)=e^{itX}$, and $\left|\frac{d}{dt}F(t,X)\right|\leq |X|$, $\left|\frac{d^2}{dt^2}F(t,X)\right|\leq |X|^2$ (we apply this result twice)

We conclude by Taylor's formula.

~

**Proof of the central limit theorem** Up to replacing $X_i$ with $X_i-\mathbb{E}[X_i]$ we can assume $\mathbb{E}[X_1]=0$.

We use Lévy's theorem. By the lemma, write $\varphi_{X_1}(t) = 1-\frac{\sigma^2}{2}t^2+\varepsilon(t)\cdot t^2$ with $\varepsilon(t)\xrightarrow[t\to 0]{} 0$.

But $\mathbb{E}[\exp(it\cdot\frac{X_1+\dots+X_n}{\sigma\sqrt{n}})] = \left(\varphi_{X_1}(\frac{t}{\sigma\sqrt{n}})\right)^n$ by $\perp\!\!\!\perp$.

Using $|u^n-v^n|\leq n|u-v|\max(|u|,|v|)^{n-1}$ for all $u,v\in\mathbb{C}$, write

$$\left|\left(1-\frac{t^2}{2n}\right)^n - \varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)^n\right| \leq n\frac{t^2}{n\sigma^2}\left|\varepsilon\left(\frac{t}{\sqrt{n\sigma^2}}\right)\right| \xrightarrow[n\to\infty]{} 0$$

In addition, $\left(1-\frac{t^2}{2n}\right)^n = \exp\left(n\ln\left(1-\frac{t^2}{2n}\right)\right) = \exp\left(n\left(-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right)\right) \xrightarrow[n\to\infty]{} e^{-t^2/2}$

We conclude that $\varphi_{\frac{X_1+\dots+X_n}{\sigma\sqrt{n}}}$ converges pointwise to the characteristic function of a $\mathcal{N}(0,1)$ r.v, this completes the proof.

~

# END OF LECTURE 26

# 6) Gaussian vectors and the multidimensional CLT (optional) (NOT part of the exam)

The extension of the Central Limit Theorem to $\mathbb{R}^d$ involves Gaussian vectors, which are also very useful in statistical models or in the study of Brownian motion

**Def** A r.v. $X=(X_1,\dots,X_d)$ with values in $\mathbb{R}^d$ is a Gaussian vector if any linear combination ($\lambda_1 X_1+\dots+\lambda_d X_d$ with $\lambda_1,\dots,\lambda_d\in\mathbb{R}$) is Gaussian, i.e follows a law $\mathcal{N}(m,\sigma^2)$ with $m\in\mathbb{R}$ and $\sigma^2\geq 0$ (by convention $\mathcal{N}(m,0)$ is a constant r.v. equal to $m$)

Recall that if $X\sim\mathcal{N}(m,\sigma^2)$ we have $\mathbb{E}[e^{iuX}]=e^{ium-\frac{\sigma^2}{2}u}$ for $u\in\mathbb{R}$

**Example** If $X_1,\dots,X_d$ are $\perp\!\!\!\perp$ Gaussian r.v, then $(X_1,\dots,X_d)$ is a Gaussian vector (we have seen that a sum of $\perp\!\!\!\perp$ Gaussian r.v. is Gaussian)

**Remark** If $(X_1,\dots,X_d)$ is a Gaussian vector, then $X_1,\dots,X_d$ are Gaussian, but the converse is false: Take $X\sim\mathcal{N}(0,1)$, if $\mathbb{P}(\varepsilon=1)=\mathbb{P}(\varepsilon=-1)=\frac{1}{2}$ and $X\perp\!\!\!\perp\varepsilon$, then $\varepsilon X\sim\mathcal{N}(0,1)$ but $(X,\varepsilon X)$ is not Gaussian since $\mathbb{P}(\varepsilon X+X=0)=\frac{1}{2}$.

**Definition** Let $X=(X_1,\dots,X_d)$ be a Gaussian vector. The vector $m_X=(\mathbb{E}[X_1],\dots,\mathbb{E}[X_d])$ is called the mean of $X$. The $d\times d$ matrix $K_X=(\mathbb{E}[X_i X_j]-\mathbb{E}[X_i]\mathbb{E}[X_j])_{1\leq i,j\leq d}$ is called the covariance matrix of $X$. $X$ is said to be centered if $m_X=(0,\dots,0)$

It is clear that $X-m_X$ is centered.

We write $\langle\cdot|\cdot\rangle$ for the scalar product on $\mathbb{R}^d$

**Proposition** Let $X$ be a Gaussian vector in $\mathbb{R}^d$ and $\lambda\in\mathbb{R}^d$. Then $\langle\lambda|X\rangle$ is a $\mathcal{N}(m_\lambda,\sigma_\lambda^2)$ r.v. with $m_\lambda=\langle\lambda|m_X\rangle$ and $\sigma_\lambda^2=\langle\lambda|K_X\lambda\rangle$.

In particular, $K_X$ is a symmetric $\geq 0$ matrix.

**Proof** • The fact that $m_\lambda = \langle \lambda | m_X \rangle$ follows from linearity of expectation

• $\mathbb{E}[(\langle \lambda | X \rangle)^2] = \sum_{i,j} \lambda_i \lambda_j \mathbb{E}[X_i X_j]$ and $\langle \lambda | m \rangle^2 = \sum_{i,j} \lambda_i \lambda_j \mathbb{E}[X_i]\mathbb{E}[X_j]$

Thus $\mathrm{Cov}(\langle \lambda | X \rangle) = \sum_{i,j} \lambda_i (K_X)_{i,j} \lambda_j = \langle \lambda | K_X \lambda \rangle$.

**Corollary** The characteristic function of a Gaussian vector $X$ in $\mathbb{R}^d$ is given by

$$\Phi_X(\lambda) = \exp\left(i \langle \lambda | m_X \rangle - \frac{1}{2} \langle \lambda | K_X \lambda \rangle\right)$$

In particular, the law of a Gaussian vector is determined by its mean and covariance matrix

**Proof** We have seen that for $\lambda \in \mathbb{R}^d$, $\langle \lambda | X \rangle \sim \mathcal{N}(\langle \lambda | m_X \rangle, \langle \lambda | K_X \lambda \rangle)$, so for every $u \in \mathbb{R}$, $\mathbb{E}[\exp(iu\langle \lambda | X \rangle)] = \exp\left(iu\langle \lambda | m_X \rangle - \frac{u^2}{2} \langle \lambda | K_X \lambda \rangle\right)$

We get the result by taking $u = 1$.

**Application** If $X, Y$ are $\perp\!\!\!\perp$ Gaussian vectors in $\mathbb{R}^d$, then $X+Y$ is a Gaussian vector with $m_{X+Y} = m_X + m_Y$ and $K_{X+Y} = K_X + K_Y$.

**Proposition** Let $m \in \mathbb{R}^d$ and $K$ be a symmetric $d \times d$ $\geq 0$ matrix. There exists a Gaussian vector in $\mathbb{R}^d$ with mean $m$ and covariance matrix $K$.

**Proof**: The idea is to write $K = C^2$ with $C$ symmetric (to see this, diagonalize $K$ in an orthogonal basis: $K = PDP^{-1}$ with $P$ orthogonal (${}^tPP = P{}^tP = I_d$) and $D = \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_d \end{pmatrix}$ with $a_1, \dots, a_d \geq 0$.

We then take $C = P\sqrt{D}P^{-1}$)

Then let $N = (N_1, \dots, N_d)$ be made of iid $\mathcal{N}(0,1)$ r.v. Set $X = C \cdot N + m$, which is a Gaussian vector. We compute $\Phi_X(\lambda)$ for $\lambda \in \mathbb{R}^d$. $\langle \lambda | X \rangle$ is a Gaussian r.v with mean $\langle \lambda | m \rangle$ and variance $\mathbb{E}[(\langle \lambda | CN \rangle)^2] = \mathbb{E}[{}^t\lambda CN \cdot {}^tN {}^tC \lambda] = {}^t\lambda C \underbrace{\mathbb{E}[N {}^tN]}_{I_d} {}^tC \lambda = {}^t\lambda K \lambda = \langle \lambda | K \lambda \rangle$

This shows the result.

We denote by $\mathcal{N}(m,K)$ a r.v. whose law is that of a Gaussian vector of mean $m$ and covariance matrix $K$.

**Corollary** A Gaussian vector $\mathcal{N}(m,K)$ has a density iff $\det(K_X) > 0$ and then its density is

$$\frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det K}} \exp\left(-\frac{1}{2} \langle x-m \mid K_X^{-1}(x-m)\rangle\right) dx$$

idea of the proof: change of variables in the previous construction $X = C\mathcal{N} + m$.

A very useful feature for Gaussian vectors is a simple way to check $\perp\!\!\!\perp$:

**Theorem**

① Let $X=(X_1,\dots,X_d)$ be a Gaussian vector in $\mathbb{R}^d$. Then $X_1,\dots,X_d$ are $\perp\!\!\!\perp$ iff $K_X$ is a diagonal matrix (i.e. $\mathrm{Cov}(X_i,X_j)=0$ for $i\neq j$)

② Let $Z=(X_1,\dots,X_p,Y_1,\dots,Y_q)$ be a Gaussian vector in $\mathbb{R}^{p+q}$. Then $(X_1,\dots,X_p) \perp\!\!\!\perp (Y_1,\dots,Y_q)$ iff $\mathrm{Cov}(X_i,Y_j)=0$ for every $1\le i\le p$, $1\le j\le q$.

**Proof** Without loss of generality, assume that all the r.v. are centered.

① $\boxed{\Rightarrow}$ We know that $X_i \perp\!\!\!\perp X_j \Rightarrow \mathbb{E}[X_iX_j]=\mathbb{E}[X_i]\mathbb{E}[X_j]$ which is $\mathrm{Cov}(X_i,X_j)=0$

$\boxed{\Leftarrow}$ Let $u=(u_1,\dots,u_d)\in\mathbb{R}^d$, then writing $K_X = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{pmatrix}$, we have

$\mathbb{E}[\exp(i\langle u \mid X\rangle)] = \exp(-\langle u \mid K_X u\rangle) = \prod_{i=1}^{d} \exp\left(-\frac{u_i^2\sigma_i^2}{2}\right)$ which implies that $X_1,\dots,X_d$ are $\perp\!\!\!\perp$ and $X_i \sim \mathcal{N}(0,\sigma_i^2)$

② The proof is similar

$\boxed{\Rightarrow}$ For $1\le i\le p$, $1\le j\le q$, $X_i \perp\!\!\!\perp Y_j \Rightarrow \mathrm{Cov}(X_i,Y_j)=0$

$\boxed{\Leftarrow}$ Set $X=(X_1,\dots,X_p)$ and $Y=(Y_1,\dots,Y_q)$, $u=(u_1,\dots,u_p)$, $v=(v_1,\dots,v_q)$.

The covariance matrix of $Z$ is of the form $K_Z = \begin{pmatrix} K_X & 0 \\ 0 & K_Y \end{pmatrix}$

Thus for $z=(u_1,\dots,u_p,v_1,\dots,v_q)$:

$\mathbb{E}[\exp(i\langle z \mid Z\rangle)] = \exp(-\frac{1}{2}\langle z \mid K_Z z\rangle) = \exp(-\frac{1}{2}\langle u \mid K_X u\rangle - \frac{1}{2}\langle v \mid K_Y v\rangle)$

$= \mathbb{E}[\exp(i\langle u \mid X\rangle)]\,\mathbb{E}[\exp(i\langle v \mid Y\rangle)]$

which implies $X \perp\!\!\!\perp Y$

⚠ In ② it is important that $(X_1, \dots, X_p, Y_1, \dots, Y_q)$ is a gaussian vector ($X$ and $Y$ gaussian vectors is not enough)

**Theorem (multidimensional CLT)** Let $(X_i)_{i \geq 1}$ be iid r.v in $\mathbb{R}^d$ with $\mathbb{E}[|X_1|^2] < \infty$ then

$$\frac{1}{\sqrt{n}}\left(X_1 + \dots + X_n - n\,\mathbb{E}[X_1]\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, K_{X_1}).$$

**Idea of the proof** It's essentially the same as in the case $d=1$: without loss of generality assume $\mathbb{E}[X_1] = 0$. Then for $u \in \mathbb{R}^d$:

$$\mathbb{E}\left[\exp\left(i \left\langle u \,\middle|\, \frac{X_1 + \dots + X_n}{\sqrt{n}} \right\rangle\right)\right] = \mathbb{E}\left[\exp\left(i \left\langle \frac{u}{\sqrt{n}} \,\middle|\, X_1 \right\rangle\right)\right]^n = \Phi_{X_1}\left(\frac{u}{\sqrt{n}}\right)^n.$$

Also by Taylor's formula, $\Phi_{X_1}\left(\frac{u}{\sqrt{n}}\right) = 1 - \frac{1}{2n}\langle u \,|\, K_{X_1} u\rangle + o\left(\frac{1}{n}\right)$

We conclude that

$$\mathbb{E}\left[\exp\left(i \left\langle u \,\middle|\, \frac{X_1 + \dots + X_n}{\sqrt{n}} \right\rangle\right)\right] \xrightarrow[n\to\infty]{} \exp\left(-\frac{1}{2}\langle u \,|\, K_{X_1} u\rangle\right)$$

and the result follows from Lévy's theorem.